SAS/ETS® PANEL Procedure for Panel Data

Modeling Combined Time Series and Cross-Sectional Data

Overview

Panel data occur when a panel of individuals—people, households, corporations, or otherwise—are observed over a period of time during which several observations per individual are obtained. Panel data have two dimensions: the individual dimension (or cross section) and the time dimension. The panel design for collecting data is among the most popular in econometrics for one primary reason: Because individuals are repeatedly measured, structural changes can be directly estimated while controlling for all time-constant aspects of the individuals, both measured and unmeasured. Put simply, each individual acts as its own control group.

Details

The PANEL procedure is a powerful tool for fitting linear regression models to panel data. Formally, for a panel of $N$ individuals, consider the linear regression model

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_{it} + v_i + \epsilon_{it}$$

where $i$ denotes the individual and $t$ is any one of $T$ time points. The regression model has two sets of explanatory variables: a set of $X$ variables that vary over time and a set of $Z$ variables that do not vary over time. The $v_i$ are known as individual (or cross-sectional) effects, and the $\epsilon_{it}$ are the observation-level regression errors.

PROC PANEL provides several ways to fit the preceding regression model, and each strategy differs in what it is willing to assume about the explanatory variables, the individual effects, the observation-level errors, and their relationships.

Example: Customer Loyalty Data from a Grocery Chain

This example illustrates how you can use the PANEL procedure to analyze some customer loyalty data. The data for this example are from 330 households who shopped regularly at a grocery chain in the Raleigh, North Carolina, area. The data track monthly meat expenditures for the year 2011. There are 12 monthly observations per household; some observations are missing because the household did not visit the chain during that month or did not use its loyalty card.

You want to determine the association between government assistance and meat purchases while controlling for other available factors. You fit regression models of the form

$$Meat_{it} = \beta_0 + \beta_1 Govt_{it} + \beta_2 Hsize_{it} + \beta_3 Rural_{it} + \beta_4 Alcohol_{it} + \beta_5 MealsOut_{it} + v_i + \epsilon_{it}$$

for household $i$ during month $t$. The dependent variable $Meat_{it}$ is the meat expenditure for the household during that month, and the regression variables are as follows:

- $Govt_{it}$: whether government assistance was received
- $Hsize_{it}$: household size
- $Rural_{it}$: whether the household is rural
- $Alcohol_{it}$: whether more than 10% of expenditures were on alcohol
- $MealsOut_{it}$: number of meals per week outside of household, as reported on a survey

You can use the PANEL procedure to fit more than a dozen varieties of this model, each having different assumptions. A good general starting point is the ultra-efficient random-effects estimator that you obtain by using the following statements:

```sas
proc panel data = grocery;
   id HouseID Month;
   model Meat = Govt Hsize Rural Alcohol MealsOut / ranone;
run;
```

You use an ID statement to specify both the individual variable (HouseID) and the time variable (Month). The `ranone` option specifies one-way random effects.

Figure 1 shows the parameter estimates from this model. Based on this estimator, controlling for factors such as household size and rural store location, you would conclude that households on government assistance purchase about $5.05 more in meat products per visit.
Besides parameter estimates, PROC PANEL also offers a slew of model specification tests and diagnostics. For example, Figure 2 shows a panel of residual diagnostic plots. From a residual perspective, it seems that the random effects model is well-behaved.

PROC PANEL supports more than a dozen model specification tests, including tests that detect serial correlation, poolability of cross sections, unobserved effects, and endogenous regressors. One example of testing for endogeneity is the Hausman specification test, demonstrated in Figure 3. The significant p-value means that the regressors are likely correlated with the individual-level errors, making the random-effects results unreliable.

The good news, however, is that PROC PANEL offers many alternatives to the random-effects estimator when it is shown to be inappropriate for the data at hand.

**Figure 3: PROC PANEL Hausman Specification Test**

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>DF</th>
<th>m Value</th>
<th>Pr &gt; m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3</td>
<td>25.72</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Possible alternatives to the random-effects estimator include the following:
- Fixed-effects estimator
- Hybrid estimators such as Hausman-Taylor and Amemiya-MacCurdy
- DaSilva moving average estimator
- Parks autoregressive estimator
- Dynamic panel estimators via GMM
- General GMM estimation
- Two-way fixed and random effects

You can obtain many model estimators at one time, either by using multiple MODEL statements or by specifying multiple options in a single MODEL statement.

When fitting multiple models, you can use a COMPARE statement to create convenient and customizable side-by-side comparisons of all fitted models; see Figure 4 for an example.

**Figure 4: PROC PANEL Model Comparison Table**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1 RunOne</th>
<th>Model 2 FixOne</th>
<th>Model 3 BtwGrps</th>
<th>Model 4 HTaylor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20.50606</td>
<td>53.894415</td>
<td>16.984418</td>
<td>19.125895</td>
</tr>
<tr>
<td></td>
<td>2.33269</td>
<td>1.549962</td>
<td>1.700415</td>
<td>2.403772</td>
</tr>
<tr>
<td>Govt</td>
<td>5.050562</td>
<td>3.591205</td>
<td>13.400587</td>
<td>3.583391</td>
</tr>
<tr>
<td></td>
<td>0.598942</td>
<td>0.655025</td>
<td>0.908573</td>
<td>0.664876</td>
</tr>
<tr>
<td>Hsize</td>
<td>5.145648</td>
<td>0.477447</td>
<td>5.092447</td>
<td>5.173890</td>
</tr>
<tr>
<td></td>
<td>0.474892</td>
<td>0.303155</td>
<td>0.352344</td>
<td>0.357340</td>
</tr>
<tr>
<td>Rural</td>
<td>-1.410680</td>
<td>-1.454439</td>
<td>1.005439</td>
<td>-1.439905</td>
</tr>
<tr>
<td></td>
<td>0.344892</td>
<td>0.357766</td>
<td>1.403805</td>
<td>0.357340</td>
</tr>
<tr>
<td>Alcohol</td>
<td>2.982397</td>
<td>2.992035</td>
<td>1.082457</td>
<td>2.974996</td>
</tr>
<tr>
<td></td>
<td>0.196014</td>
<td>0.201343</td>
<td>1.768137</td>
<td>0.200391</td>
</tr>
<tr>
<td>MealsOut</td>
<td>-2.827608</td>
<td>-2.676694</td>
<td>-2.679642</td>
<td>-1.922421</td>
</tr>
<tr>
<td></td>
<td>0.384842</td>
<td>0.262948</td>
<td>0.268967</td>
<td>0.808967</td>
</tr>
</tbody>
</table>

**System Requirements**

Estimation as demonstrated in these examples requires:
- Base SAS® 9.4 or later
- SAS/ETS® 14.1 or later